DOCUMENT RESUME

ED 205 895 CS 006 178

AUTHOR Klare, George R.
TITLE Readability: Passage Difficulty and Context Dependence.
PUB DATE 81
NOTE 27p.: Paper presented at the Annual Meeting of the International Reading Association (26th, New Orleans, LA, April 27-May 1, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS College Freshmen: *Difficulty Level: High School Seniors: *Prior Learning: *Readability: Readability Formulas: *Reading Comprehension: *Reading Interests: *Reading Research: Test Reliability
IDENTIFIERS Schema Theory

ABSTRACT

Two studies investigated the relationships between readability, prior knowledge, and interest in the material being read. In the first study, 169 college freshmen read cloze passages of materials from the Nelson-Denny Reading Tests, then one week later read the whole passages and completed the related multiple choice questions. A control group merely read the passages and answered the multiple choice questions. The cloze scores correlated with test passage readability, but the multiple choice scores did not correlate with the readability scores. When data from 50 more subjects who took the multiple choice tests without reading the passages were entered into the calculations much higher correlations were found between these "corrected" scores and the readability scores. In the second study, 266 entering freshmen supplied reading interest data and read high and low interest passages with manipulated reading difficulty levels. The results showed that prior knowledge and reading interest were significantly related. Further analyses of data from 18 subjects selected for combinations of interest and readability revealed that prior knowledge was the more powerful variable in the study. The collective results from the two studies indicated that readability measures tend to overestimate passage difficulty when prior knowledge and reading interest are high, but underestimate passage difficulty when these variables are low. (RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED205895

International Reading Association

26th Annual Convention, April 27--May 1, 1981, New Orleans, La.

Symposium 10--Comprehension: Context Dependence, Readability, and Comprehensibility of Subject Matter Text

Readability: Passage Difficulty and Context Dependence

George R. Klare, Ohio University

To begin, let me give you a bit of background about my topic, and my interest in it. About five years ago I reviewed, in detail, all the experimental studies I could find which related readability to comprehension. These were studies where research workers had rewritten passages to give them a more and/or less readable score, and then had measured reader comprehension of the versions. I presented a paper based on this review called "A Second Look at the Validity of Readability Formulas" (Klare, 1976). The results of my review, I said, could be expressed whimsically by the following slide.

Slide 1: Does Readability Make a Difference? Sometimes

More seriously and more specifically, the next slide gives a capsule summary covering the 36 studies I was able to locate at that time. In this slide, keep in mind that:

- "positive" means the studies yielded statistically significant differences favoring the more readable version or versions;
- "negative" means the studies yielded non-significant differences between the more and less readable versions; and
- "mixed" means that some results were significant and some were not.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
George R. Klare
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

CS206178

Slide 2: 36 Experimental Studies...

As you can see, around 3/5 to 2/3 of the results were significant. The rest were not, and these non-significant results interested me as much as, if not more than, the significant ones. If you want to find out when and why something works, and when and why it does not work, you need to make comparisons. So, you need studies of both kinds.

Readability formulas have come to be widely used in making educational decisions, and we need therefore to know the limits of their usefulness: when they work, and when they can be misleading. I might just add one thing about the studies with non-significant results: there are no such studies in the articles column. Thus the slide appears to bear out the common wisdom that journal editors, especially, do not like to publish non-significant results.

Actually, however, the reason for that appearance is something different. When a study appeared in both dissertation or published form, I chose the dissertation. I found that dissertations have far more details and are thus more useful than articles. Graduate students cannot hide their mistakes as writers of articles can, because of the length limitations of journal articles.

I examined these 36 studies in great detail, starting with 40 characteristics I felt might be important. I could not find all of these covered in all of the studies, but did end up with 28 characteristics, under the following general categories.

1. The experimental passages, and how they were modified.
2. The tests and other dependent measures used.

Slide 2

36 Experimental Studies of the Effect of Readability Variables Upon Comprehension

Results	All Studies	Published Studies	Theses or Dissertations
Positive	19	6	13
Mixed	6	3	3
Negative	11	0	11
	36	9	27

3. Descriptions of the subjects and their characteristics.

4. The instructions given to the subjects.

5. The statistical analyses employed.

6. The results, and detailed discussions based on them.

Out of my analysis, I developed a framework, or model, of some major factors interacting with readability measures in validity studies. The next slide presents this model.

Slide 3: Some Major Factors...

Certain boxes in the model can be thought of as "contexts." Whether formulas are likely to give reliable predictions depends to some extent upon the contexts in which they are used. This may not sound surprising, but when you think about it for a minute, you realize that formulas appear to be relatively "context-free." They appear, in other words, to be equally useful in all contexts in which text, or prose, is used. I use the word "relatively" because we would not assume they would be useful with everything, for example, such materials as:

·mathematical formulas; or

·poetry; or

·some language other than the one for which they were developed.

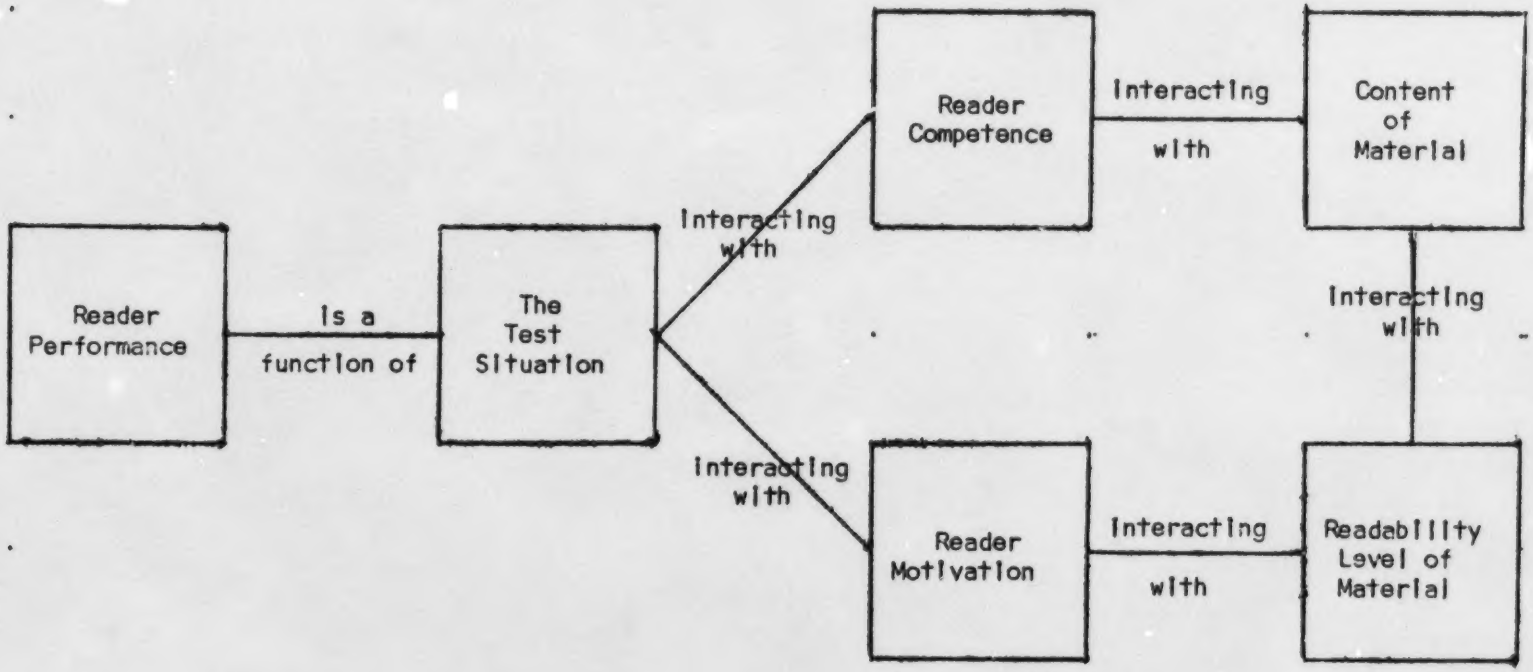
But we might assume they would be equally useful for different subject-matter areas, different characteristics of students, or different conditions of use. After all, this is one of the strengths of formulas, that they are relatively context-free.

The non-significant results suggested, however, that there were some

Slide 3

G. R. Klare

Some Major Factors Interacting with
Readability Measures in Validity Studies



contexts where formulas might not give reliable predictions. The various boxes in the model are actually summary statements of the major factors which, my analysis suggested, interact with each other in producing some particular performance. Each box contains a number of variables in a finer breakdown. I will not present this breakdown except for the one box most relevant to our symposium--that concerned with "Content of Material." The next slide shows what is in this box.

Slide 4: Some Content Considerations

I would like to present the results of several studies we have done recently covering two of the variables listed in the box:

- new information (in relation to reader knowledge)--now often called "prior knowledge," and
- interest-value (in relation to reader interests).

I also need to say, at the outset, that these studies did not involve K to 12 students, as the description of the symposium suggests. All of them involved college freshmen. But since many of the students were tested before they entered college (during their summer orientation), perhaps they still qualify as high school seniors. I believe the results are general enough to apply to many other settings, as you will see.

First, let me describe some work with the first variable, "new" vs. "old" information, or Prior Knowledge. We will look at its relationships to readability and comprehension. This work was done with Eileen Entin, a former graduate student of mine who has now completed her doctoral work at Ohio University.

Slide 4

Some Content Considerations

A. New information (in relation to reader knowledge)

B. Interest-value (in relation to reader interests)

C. Nature of content (in relation to reader intellectual level)

D. Maturity of content (in relation to reader maturity)

The first study was published under the title of "Some Interrelationships of Readability, Cloze, and Multiple-Choice Scores on a Reading Comprehension Test" (Entin and Klare, 1978). We chose the Nelson-Denny Reading Test for our work, and in particular the "Comprehension Test and Reading Rate" section. This section has eight passages, seven with approximately 200 words each, and one with about 400 words. For this study, we used the seven 200-word passages; the 400-word passage was used only as a control and covariance measure, and we can forget it for the moment.

We administered the seven passages to college freshmen to compare their scores on the tests' multiple-choice questions with their scores on cloze versions of the passages. We handled the testing this way.

1. During the first week we gave 79 subjects the regular cloze version of the seven passages. Most of you are probably familiar with cloze tests, but just in case, I will put an example on the screen. This regular version will be referred to as "Cloze 1."

Slide 5: Standard-Blank Cloze Form

---

2. We were, at the time, experimenting with a new type of cloze, which I will call dash-blank, or "Cloze 2." So we used it also, in this case on 90 subjects. Here is what it looks like.

Slide 6: Dash-Blank Cloze Form

---

3. One week later, we gave both groups of subjects the seven Nelson-Denny passages in their complete form--that is, without any blanks. We then had them take the regular multiple-choice questions published for use with the passages.

Slide 5

MC-D
53

G. R. Klare

Long before _____ (1) days of printing, minstrels _____ (2)
from castle to castle _____ (3) before kings and their _____ (4).
Their songs were usually _____ (5) the character and the _____ (6)
deeds of a real _____ (7). Often these minstrels, these _____ (8)
used their imagination and _____ (9) mystical deeds. No one
_____ (10) the time attempted to _____ (11) down any of these
_____ (12), for few knew how _____ (13) write. The stories were
_____ (14) handed down by word _____ (15) mouth, very much in
_____ (16) same way as were _____ (17) legends of the American
_____ (18). But about A.D. 700 _____ (19) stories relating to the
_____ (20) deeds of a hero, _____ (21) were collected by some
_____ (22) poet of the time. _____ (23) poem has since been
_____ (24) into modern English and _____ (25) we can enjoy reading
_____ (26) first epic poem in _____ (27) literature. Beowulf fought
two _____ (28) fights to save a _____ (29) and one to save
_____ (30) own people.

Slide 6

MC-D
53

G. R. Klare

Long before the days __ printing, minstrels wandered from _____ to castle (1) (2)

singing before _____ and their retainers. Their _____ were usually about (3) (4)

the _____ and the brave deeds __ a real hero. Often _____ (5) (6) (7)

minstrels, these "gleemen," used _____ imagination and added mystical _____. (8) (9)

No one at the _____ attempted to write down ___ of these tales, for ___ knew (10) (11) (12)

how to write. ___ stories were originally handed _____ by word of mouth, (13) (14)

_____ much in the same ___ as were the legends __ the American Indians. But (15) (16) (17)

_____ A. D. 700 the stories _____ to the brave deeds __ a hero, (18) (19) (20)

Browulf, were _____ by some Anglo-Saxon poet __ the time. This poem (21) (22)

___ since been translated into _____ English and today we ___ enjoy (23) (24) (25)

reading the first _____ poem in English literature. _____ fought two (26) (27)

dreadful fights __ save a king and ___ to save his _____. (28) (29) (30)

4. To see whether having the cloze test one week earlier had raised the level of information about the passages, we had a third group who did not take a cloze test the week before. The subjects simply read the regular passages and took the multiple-choice tests based on them during the second week. We found, with the help of a covariance analysis, no evidence that the multiple-choice scores of the first two groups had been affected by having the cloze test the week before.
5. We also analyzed the Flesch Reading Ease readability values for the seven passages (Flesch, 1948). We wanted to relate these values to the cloze scores and the multiple-choice scores on the passages.

Following this, we ran correlations of the above measures. Some of the results we expected, but others surprised us. The next slide presents the results. It has a lot of information on it, so I will do some interpreting after I present it.

Slide 7: Pearson Coefficients...

First, the <u>good</u> news--it is in the yellow-shaded section of the slide. Note the following.

- The mean scores on the regular type of cloze--Cloze 1--correlated .68 with the readability scores on the seven passages.
- The mean scores on the dash-blank type of cloze--Cloze 2--correlated .74 with the readability scores.
- The two types of cloze correlated .97 with each other.

Next, the <u>bad</u> news--it is in the pink-shaded section of the slide. These

Slide 7

Correlations Among Original and Corrected
Estimates of Difficulty for Seven Passages

	MC-S Uncorrected	MC-D Uncorrected	MC-S Corrected	MC-D Corrected
RE[a]	-.10	.01	.41	.45
Cloze-S[b]	-.22	-.08	.48	.53
Cloze-D[c]	-.17	-.04	.45	.50
MC-S (Uncorr.)[d]		.98**	.34	.29
MC-D (Uncorr.)[e]			.49	.45
MC-S (Corrected)[f]				.99**
MC-D (Corrected)[g]				

**$p < .001$

[a]Flesch Reading Ease (1948)
[b]Solid Line cloze score
[c]Dash Line cloze score
[d]Uncorrected multiple choice score for subjects in the solid line cloze group
[e]Uncorrected multiple choice score for subjects in the dash line cloze group
[f]Corrected multiple choice score for subjects in the solid line cloze group
[g]Corrected multiple choice score for subjects in the dash line cloze group

correlations involve the multiple-choice scores; they are labelled "Uncorr" in the slide, meaning "Uncorrected," for a reason I will explain in a minute. All we need to know at this point is that they are the scores as I described them earlier. Note the following.

- The mean multiple-choice scores on the seven passages correlated -.11 and .01 with the readability scores--in other words, essentially zero. (The reason for two scores is that there were two groups of subjects; one group had Cloze 1 the first week and the other Cloze 2 the first week.)
- The mean multiple-choice scores on the seven passages correlated from -.04 to -.22 for the two groups and the two types of cloze--again essentially zero.

This raised an interesting question: Why were there no clear relationships? Studies in the literature had led us to expect them. Our experience, as well as the Content Considerations box in the model, led us to examine next the questions for the Nelson-Denny passages. Like Farr and Smith (1970) and Tuinman (1973-74) before us, we found them less than perfect. Specifically, we felt that a number of them probably could have been answered by our subjects before reading the passages they covered.

So we collected a new sample of 50 freshmen and gave them the Nelson-Denny questions without ever having read the passages. We then adjusted the original mean scores, which we called "Uncorrected," for amount of Prior Knowledge. We did this by subtracting, for each item, the second group's mean score from the first group's mean score. These we called "Corrected" multiple-choice items, or "Corr" in the slide. The correlations involving these mean scores are shaded in blue on the slide.

Return Slide 7: Pearson Coefficients...

Notice, now, that the correlations changed markedly.

- The corrected mean multiple-choice scores correlated .41 and .44 with the readability scores.
- These corrected mean scores also correlated from .45 to .50 with the cloze scores.

Although we cannot be sure, we suspect that the surprising jumps in these correlations might even have jumped a bit higher if we had been able to use the same subjects for our correction, instead of different ones. We could not do that, but I will have another way to look at the role of Prior Knowledge in a minute.

About the time our article covering this study was in press, the Hanna-Oaster article on context dependence appeared (Hanna and Oaster, 1978-79). Since our study had been set up in a way that permitted us to make their more refined analysis of our multiple-choice results, we did so (Entin and Klare, 1980). The results appear in the next slide.

Slide 8: Correlations Among...

As you probably recall, the Hanna-Oaster abbreviations are as follows:

- CII--Context Independence Index, which refers to the mean percentage of multiple-choice questions correct before reading the passage they cover.

Slide 8

Correlations among Nelson-Denny values for *Passage* indices (N=7)

	CII	CDI	RE	Solid Cloze	Dash Cloze
CII	1.00	- .69***	- .46*	- .62**	- .55**
CDI	- .69***	1.00	.39	.46*	.40
RE	- .46*	.39	1.00	.68***	.74***
Solid Cloze	- .62**	.46*	.68***	1.00	.97***
Dash Cloze	- .55**	.40	.74***	.97***	1.00

Note: With N=7, probabilities for starred items are as follows:
*** $p < .05$
** $.05 < p \leq .10$
* $.10 < p \leq .15$

·CDI--Context Dependence Index, which refers to the mean percentage of multiple-choice questions correct after subtracting the above percentage and also the chance score.

Now, as the slide shows:

·Readability scores show a negative correlation of -.46 with the Context Independence Index (CII), but a positive correlation of .39 with the Context Dependence Index (CDI).

·The cloze scores show negative correlations of -.62 and -.55 with the Context Independence Index, but positive correlations of .46 and .40 with the Context Dependence Index.

What do these results say? As predicted by the model and the paper in which it appeared (Klare, 1976): where Prior Knowledge of content is high, the effect of Readability may be attenuated or eliminated.

As mentioned earlier, Eileen Entin and I also looked at another of the variables in the "Content" box from the model--Interest-Value. In the process, we also looked more closely at Prior Knowledge. This was in Eileen's doctoral dissertation, as yet unpublished, and was carried out in the following manner.

1. During the period of college orientation, a survey was made of freshmen's preferences for a list of 120 topics selected from the World Book Encyclopedia. Six topics of highest and six of lowest interest were chosen on the basis of the preferences of about 525 subjects. The next slide presents the twelve topics.

Slide 9: Means and Percentage Distributions...

Slide 9

Means and Percentage Distributions for Ratings of Interest in 12 Selected Topics (Preliminary Study)

Topic	n	Mean	Uninteresting	Moderately Interesting	Interesting
Properties of Aluminum Alloys (Aluminum)	525	2.75	79.8	15.6	4.6
How Cement is Made (Cement)	524	2.72	75.8	20.4	3.8
Basic Properties of Crop Production (Crops)	519	2.66	72.6	20.8	6.6
How Clothes are Dry Cleaned (Dry Clean)	521	2.72	77.5	16.9	5.6
Processes in Meat Packing (Meat)	514	2.69	73.3	22.4	4.3
Rocks as a Hobby (Rocks)	502	2.71	74.9	20.7	4.4
Attitudes About Death (Death)	524	1.62	14.5	32.6	52.9
How Emotions Help and Harm Us (Emotions)	519	1.54	11.9	29.7	58.4
Why We Forget (Forget)	490	1.60	14.3	31.4	54.3
The Nature of Personality (Personality)	500	1.57	14.6	27.8	57.6
Keeping Physically Fit (Physical Fit)	506	1.55	10.7	33.4	55.9
What Happens While We Sleep (Sleep)	495	1.56	12.3	31.1	56.6

2. Passages of approximately 400 words corresponding to these topics were taken from the Encyclopedia and rewritten to "Standard" and "Difficult" levels according to the Flesch Reading Ease readability formula. These levels correspond, approximately, to high school graduate and college graduate reading levels according to Flesch's interpretation table. Judges verified that the content of the two versions of the passages was the same.
3. Four measures of Prior Knowledge were obtained:
   a. three subjective ratings, based on judgments before and after testing for comprehension; and
   b. one objective measure, derived from subjects' definitions of the key terms used in the passages, and administered before testing for comprehension.
4. Cloze tests on the twelve passages provided the measure of comprehension. The typical 1:5 deletion ratio was used, and two of the five possible forms were administered. This was done to avoid the problem of getting only an unusually easy or difficult single form by chance. The first 250 words of the passages were left unchanged, with the cloze tests based on the last 150 words (i.e., 30 blanks). These tests were administered to 266 freshmen.

The results of the study were as follows. First, the ratings of Interest and Prior Knowledge were related--in fact, for all topics the correlations were significant beyond the .001 level. This meant that it was easy to find topics and passages for which both Interest and Prior Knowledge were either high or low. But it also meant that it was very difficult to find topics and passages for which one was high and the other was low. This complicated

matters, since the planned within-subjects analysis of variance involving Prior Knowledge became impractical. We were, however, able to do the planned analysis involving only the Interest and Readability variables. This analysis is presented in the next slide.

Slide 10: Interest by Readability...

Note that the analysis on the left side of the slide involves all subjects, while that on the right side is for a subsample. The subsample consisted of subjects whose scores fell within one grade level below and one grade level above 12 on the Nelson-Denny Reading Test. The purpose of doing this was to keep the mean Nelson-Denny score the same but reduce the standard deviation. This meant that the "Standard" passage would be more like the tested reading level of the subsample, and the "Difficult" passage would be more clearly above their tested reading level. (This was designed, in other words, to serve the same purpose as setting up confidence intervals.)

As indicated, both Interest and Readability were highly significant in both analyses. Passages at higher levels of Interest produced higher cloze scores than those at lower Interest levels. Also, passages at the Standard level of Readability produced higher cloze scores than passages at the Difficult level.

We had hoped, as I said, to do a within-subjects analysis of variance involving Prior Knowledge. Because of the relationship between Interest and Prior Knowledge, however, we had trouble finding topics high in one but low in the other characteristic, and vice versa. Consequently, we had to use the objective measure of Prior Knowledge as a covariate. We could then do a

Slide 10

Interest by Readability Analysis of Variance

(Within Subjects) on Comprehension Scores

	Total Sample (n=243)		Subsample (n=40)	
	Table of Means		Table of Means	
	Readability (R)		Readability (R)	
	Standard	Difficult	Standard	Difficult
Interest (I)				
Low	25.83	21.70	25.97	22.30
High	29.33	25.86	29.38	25.17

ANOVA Table

Source	MS	df	F	p
I	3559.17	1	199.17	.0001
error	17.87	242		
R	3498.14	1	165.40	.0001
error	21.15	242		
I x R	27.00	1	1.43	NS
error	18.87	242		

ANOVA Table

Source	MS	df	F	p
I	393.76	1	19.86	.0001
error	19.83	39		
R	620.15	1	24.65	.0001
error	25.16	39		
I x R	2.76	1	.12	NS
error	23.29	39		

within-subjects analysis of covariance involving Interest and Readability. The results of this analysis are presented in the next slide.

Slide 11: Interest by Readability...

Note, again, that the left side of the slide presents the analysis for the entire sample, and the right side for the subsample. The effect of Interest is greatly diminished in the total-subjects analysis, and is not significant in the subsample analysis. It becomes even clearer here that Interest and Prior Knowledge involve, to some extent at least, the same variance.

So we seem to have a chicken-egg problem: Which comes first, Interest or Prior Knowledge? Expressed that way, we might be tempted to say that if something is interesting, we learn more about it. But we also might say that some topics become interesting as our knowledge increases.

Fortunately, we did not have to resolve that controversy in the dissertation. All we could hope to decide was whether Interest or Prior Knowledge contributed more of the variance. We could get at least a partial answer if we could find particular subjects who had either high or low Prior Knowledge scores at each combination of Interest and Readability levels. This would make between-subjects analysis of variance possible. We found eighteen such subjects, and the analysis was run. The results are presented in the next slide.

Slide 12: Interest by Readability...

Slide 11

Interest by Readability Analysis of Covariance

(Within Subjects) on Comprehension Scores

	Total Sample (n=204)		Subsample (n=36)	
	Table of Adjusted Means		Table of Adjusted Means	
	Readability (R)		Readability (R)	
	Standard	Difficult	Standard	Difficult
Interest (I)				
Low	26.13	22.14	27.35	23.30
High	28.60	25.30	27.51	24.34

ANCOVA Table

Source	MS	df	F	p
I	261.20	1	14.46	.0001
Cov.	40.81	3	2.26	.08
error	18.06	200		
R	2737.11	1	134.83	.0001
Cov.	23.10	3	1.14	NS
error	20.30	200		
I x R	21.88	1	1.19	NS
Cov.	40.86	3	2.22	.09
error	18.40	200		

ANCOVA Table

Source	MS	df	F	p
I	40.27	1	1.38	NS
Cov.	3.67	3	.17	NS
error	21.44	32		
R	488.55	1	20.16	.0001
Cov.	23.05	3	.95	NS
error	24.23	32		
I x R	10.32	1	.50	NS
Cov.	30.39	3	1.46	NS
error	20.81	32		

Slide 12

Interest by Readability by Prior Knowledge Scores

Analysis of Variance on Comprehension Scores

Table of Means

	Low Prior Knowledge (PK) (n=9)		High Prior Knowledge (n=9)	
	Readability (R)		Readability	
Interest (I)	Standard	Difficult	Standard	Difficult
Low	12.56	10.28	14.78	13.44
High	12.94	10.94	16.83	14.28

ANOVA Table

Source	MS	df	F	p
Prior Knowledge	178.92	1	8.43	.01
error	21.23	16		
Interest	17.50	1	3.39	.08
I x PK	3.78	1	.73	NS
error	5.13	16		
Readability	75.03	1	14.63	.001
R x PK	.17	1	.03	NS
error	5.13	16		
I x R	1.00	1	.14	NS
I x R x PK	2.53	1	.35	NS
error	7.14	16		

In this analysis, Prior Knowledge was significant at the .01 level and Readability at the .001 level, but Interest reached only the .08 level. So, whether or not Prior Knowledge came first, it appeared at least to be the more powerful variable in our study. This does not, of course, say Interest is unimportant. The two variables do not overlap completely, and it may well be that if more subjects had been available all three variables, Prior Knowledge, Interest, and Readability, would have been significant contributors.

What, then, does this set of studies say about the contexts which may affect passage difficulty, or comprehension? The serious contenders must include degree of Prior Knowledge and of Interest-Value. In other words, readability measures may tend to:

- overestimate passage difficulty when the degrees of these variables are high, and
- underestimate passage difficulty when the degrees of these variables are low.

That is why, you see, my review of the literature on studies where readability was changed led me to say that:

- Readability Makes a Difference--Sometimes.

When depends, to some extent at least, on the context in which it is applied.

References

Entin, E. B. Relationships of measures of interest, prior knowledge, and readability to comprehension of expository passages. Unpublished doctoral dissertation, Ohio University, 1980.

Entin, E. B., and Klare, G. R. Some inter-relationships of readability, cloze, and multiple-choice scores on a reading comprehension test. Journal of Reading Behavior, 1978, 10, 417-436.

Entin, E. B., and Klare, G. R. Components of answers to multiple-choice questions on a reading comprehension test: An application of the Hanna-Oaster approach. Reading Research Quarterly, 1980, 15, 228-236.

Farr, R., and Smith, C. B. The effects of test item validity on total test reliability and validity. In G. Schick and M. M. May (Eds.), Reading: Process and pedagogy. Nineteenth Yearbook of the National Reading Conference, Vol. 1. Milwaukee, Wisconsin: National Reading Conference, 1970, 122-134.

Flesch, R. F. A new readability yardstick. Journal of Applied Psychology, 1948, 32, 221-233.

Hanna, G. S., and Oaster, T. R. Toward a unified theory of context dependence. Reading Research Quarterly, 1978-79, 14, 226-243.

Klare, G. R. A second look at the validity of readability formulas. Journal of Reading Behavior, 1976, 8, 129-152.

Tuinman, J. J. Determining the passage dependency of comprehension questions in five major tests. Reading Research Quarterly, 1973-74, 9, 206-223.